

Introducció a l'estadística dels videojocs

Autor: Àlex Porcel Julià
Tutora: Alejandra Cabaña



Universitat Autònoma de Barcelona

Abstract

Català:

La indústria del videojoc ha experimentat una evolució enorme en les últimes dècades, tant en els dispositius que s'utilitzen per a jugar com en els pròpys videojocs. Desde les primeres màquines recreatives, fins als jocs en realitat virtual que comencen a sortir al mercat recentment, les empreses que formen aquesta indústria han hagut d'adaptar-se a una gran quantitat de canvis en poc temps. Un dels canvis més recents i radicals alhora ha sigut l'aparició dels dispositius intel·ligents, que han transformat completament la forma de dissenyar videojocs. Aquest fet ha causat la creació d'una branca de la ciència on es combinen l'estadística i la informàtica, que s'utilitza per estudiar el rendiment d'un videojoc, anomenada *game analytics*.

Aquest estudi consta d'una introducció a l'estadística en el món dels videojocs, i d'una simulació i el posterior anàlisi de les dades d'un tipus concret de videojoc, els jocs gratuïts per a dispositius intel·ligents, passant per tots els aspectes principals que ha d'estudiar una empresa de videojocs si vol millorar la rendibilitat d'un dels seus productes.

English:

The videogames industry has suffered a huge evolution in the last few decades, on the devices used to play the games and on the videogames themselves. Since the first arcade machines to the virtual reality games that are recently being published, the companies that involves this industry have had to adapt to many changes in a small period of time. One of the most recent and radical changes has been the emergence of the smartphones, which have completely transformed the way of designing videogames. This fact has caused the creation of a new science branch, which combines statistics and computer science, used to study the performance of a videogame, the game analytics.

This study consists of an introduction to statistics in the world of gaming, a simulation and subsequent analysis of the data of a certain type of game, the free to play games for smartphones, and all the main aspects that everyone who wants to improve their game's performance should study.

Castellano:

La industria del videojuego ha experimentado una evolución enorme las últimas décadas, tanto en los dispositivos que se utilizan como en los propios videojuegos. Des de las primeras máquinas recreativas hasta los juegos en realidad virtual que empiezan a salir al mercado recientemente, las empresas que forman esta industria se han tenido que adaptar a una gran cantidad de cambios en poco tiempo. Uno de los cambios más recientes y radicales ha sido la aparición de los dispositivos inteligentes, que han transformado completamente la manera de diseñar videojuegos. Este hecho ha causado la creación de una nueva rama de la ciencia, que combina la estadística y la informática, que se utiliza para estudiar el rendimiento de un videojuego, llamada *game analytics*.

Este estudio consta de una introducción a la estadística aplicada a los videojuegos, y de una simulación y el posterior análisis de los datos de un tipo de videojuego, los juegos gratis para dispositivos inteligentes, pasando por todos los aspectos principales que tiene que estudiar una empresa que quiera mejorar la rentabilidad de sus productos.

Keywords: *free to play (F2P), key performance indicator (KPI), in-app purchase (IAP), unique users, retention, revenue.*

Índex

1	Introducció	3
1.1	Història dels videojocs	3
1.2	Videojocs per a <i>smartphones</i>	4
2	Game analytics	6
2.1	Indicadors clau del rendiment	6
2.1.1	Ús	7
2.1.2	Monetització	8
2.2	Teorema de descomposició del <i>free to play</i>	10
2.3	Dades	11
2.3.1	Política de privacitat de dades	11
2.3.2	Simulació	12
3	Anàlisi	15
3.1	Estudi del rendiment d'un <i>free to play</i>	15
3.1.1	Mètriques d'ús	15
3.1.2	Mètriques de monetització	17
3.2	Estudi del perfil d'ús dels usuaris	19
3.3	Estudi del perfil de monetització dels usuaris	24
4	Conclusions	27
5	Bibliografia	28
6	Annexes	29

1 Introducció

1.1 Història dels videojocs

El *boom* dels videojocs va ser durant la dècada del 1970, quan es van començar a produir màquines recreatives per a locals d'oci, on l'usuari pagava per a jugar una partida o un nombre de partides determinat. L'èxit que van tenir jocs com els mítics Pong publicat per l'empresa Atari, o Space Invaders, de Taito, va fer que s'expandissin ràpidament arreu del món, principalment a Estats Units, Europa i Japó.

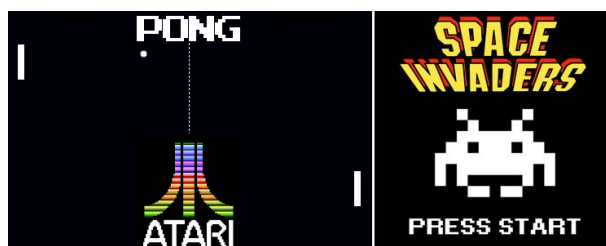


Figura 1: La primera versió del Pong d'Atari va sortir al 1972, mentre que Space Invaders es va publicar al 1978.

A partir d'aquí, va arribar el que es considera com l'edat d'or dels videojocs recreatius, protagonitzada per les millors tècniques que es van dur a terme en els dispositius i la innovació en els propis videojocs, que fins aleshores no deixaven de ser còpies dels ja existents. Amb l'aparició de jocs tant mundialment coneguts com el Pac-man, publicat per Namco, i també d'empreses com Nintendo, que va treure al mercat el Donkey Kong i va incorporar el Tetris en els seus dispositius durant la dècada dels 80, els videojocs es van establir en la cultura d'entreteniment de la societat, i desde aleshores, aquesta relació s'ha fet molt més forta.



Figura 2: El Pac-man es va publicar l'any 1980, mentre que el Tetris ho va fer juntament amb la primera Game Boy de Nintendo l'any 1989, tot i que s'havia creat cinc anys abans i extès arreu del món.

Pel que fa a la instauració dels videojocs a les llars, va ser durant la segona meitat d'aquesta mateixa dècada dels 80 que van sortir al mercat consoles com la Sega Master System o la NES (Nintendo Entertainment System) donant a conèixer per primera vegada el joc Super Mario Bros, i que també incloïa gran part dels videojocs coneguts fins al moment. Però la gran innovació d'aquestes consoles va ser el fet de que per a executar un joc s'havia d'insertar un cartutxo

que contenia el codi font a la consola, que el llegia i l'executava per a que es veiés per la pantalla. Així doncs, ja no eren com les màquines recreatives, que cada una servia per un joc determinat, si no que tenir la consola permetia executar diferents jocs, i només haver-se de comprar el cartutxo dels jocs nous a mesura que anaven sortint.



Figura 3: *Super Mario Bros* va sortir amb la NES l'any 1985, però no va ser la primera aparició d'aquest mític personatge, ja que havia estat el protagonista de *Donkey Kong*, on l'usuari havia de guiar-lo per a rescatar la princesa del gorila.

Les consoles de videojocs han continuat la seva evolució passant, tant pels diferents models de Nintendo, com per l'aparició de la PlayStation¹. Paral·lelament, els videojocs per a ordinadors personals, tant de taula com portàtils, han experimentat una evolució semblant, sobretot desde que les consoles tenen accés a la xarxa i han pogut adaptar-se a les exigències del públic: els jocs multijugador *online*.

1.2 Videojocs per a *smartphones*

Amb l'aparició dels dispositius intel·ligents que actualment tots tenim a l'abast, ja sigui mòbil, tauleta, televisió o d'altres, el sector del videojoc ha hagut de reinventar-se una vegada més per adaptar-s'hi. El model de negoci de la indústria funcionava d'una manera relativament senzilla: les empreses que es dedicaven a crear videojocs, anomenades *developers*, creaven un videojoc per a consola i/o ordinador, el posaven a la venda i els usuaris el compraven per jugar-hi en els seus dispositius. Amb l'arribada dels dispositius intel·ligents, això ha canviat i s'ha implantat un sistema que tots coneixem avui en dia que són els mercats d'aplicacions.

Els mercats d'aplicacions són botigues dins del *smartphone* en les que una empresa o persona particular hi pugen els seus productes per a que els usuaris els trobin i se'ls descarreguin en el seu dispositiu. De la mateixa manera que hi ha aplicacions i jocs que s'han de pagar, n'hi ha d'altres que són gratuïts, els *free to play* (F2P), i els usuaris hi poden jugar sense necessitat de gastar-se diners. Aquest fet incentiva a alguns usuaris a provar tot tipus de jocs, ja que només els hi costa el temps de descarregar-se'l, i si no els agrada quan hi juguen, el borren i en busquen un altre.

¹La PlayStation va sortir al mercat de Japó de la mà de Sony Computer Entertainment el 1994 i a la resta del món durant el 1995.

Existeixen dos grans mercats d'aquest tipus, que són l'*AppStore*, únicament per a dispositius amb sistema operatiu iOS (*i-Operating System*), i el *Google-Play*, per a dispositius amb sistema operatiu Android. El mercat que utilitza el dispositiu és un o l'altre, en funció de la marca fabricant, tot i que el contingut és en gran part el mateix en els dos. A part d'aquests dos, hi ha altres mercats d'aplicacions independents que es poden descarregar al dispositiu, tot i que s'utilitzen molt poc.

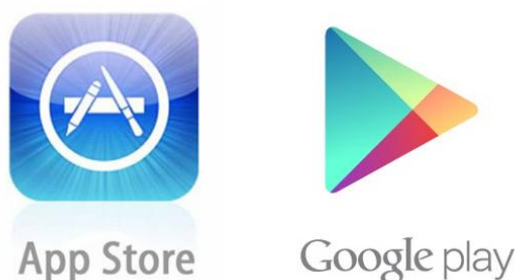


Figura 4: Icones dels mercats d'aplicacions dels dispositius amb sistema operatiu iOS (esquerra) i Android (dreta).

La pregunta que sorgeix a continuació és òbvia: com es financen les empreses que es dediquen a fer jocs *free to play*? Per a respondre-la és necessari introduir un concepte nou, que s'anomena IAP. De les sigles *In-App Purchase*, les IAP són microtransaccions que es poden efectuar dins del pròpi joc per tal d'aconseguir recompenses especials, desbloquejar contingut extra o simplement generar un estalvi en el temps de joc de l'usuari². Les IAP serveixen per finançar els videojocs F2P, que acostumen a sortir al mercat en una primera versió simplificada i depenent de la seva rendibilitat s'inverteixen recursos en incloure noves funcionalitats, o no. El que és evident és que, com que l'usuari ja no paga al comprar-se el joc, els *developers* han de mantenir-lo actualitzat i afegir-hi continguts nous constantment, i si aconsegueixen tenir un producte prou atractiu, l'abast i l'èxit que pot tenir és mundial i instantani. És clar que una gran part dels usuaris no farà cap compra si pot continuar jugant de franc, per molt atractiu que sigui el contingut de les IAP, però hi haurà un petit percentatge que sí que ho farà, i així com molts d'aquests només faran una sola compra, una petita part es poden arribar a gastar grans quantitats de diners.

Per tots aquests fets, la publicació d'un videojoc *free to play* s'ha convertit actualment en un estudi constant de l'audiència del joc, de la mateixa manera que es fa el seguiment d'un negoci qualsevol mitjançant *business intelligence*. L'estudi del comportament dels usuaris, les seves reaccions a certs canvis en el joc i sobretot, l'estudi de les vendes per a determinar si aquest joc pot ser rendible, han desembocat en l'aparició d'una nova branca d'aquest sector, anomenada *game analytics*.

²Per posar un exemple de cada un, en un joc de tirs podria desbloquejar una arma que cap usuari gratuït pot aconseguir, en un joc de plataformes podria donar accés a unes pantalles especials, i en un joc multijugador *online* podria reduir el temps que tarda en fer-se una certa millora.

2 Game analytics

Les *game analytics* són una sèrie de eines i tècniques estadístiques que s'apliquen en el sector dels videojocs per a estudiar i intentar predir el comportament dels usuaris d'un joc *free to play* per tal de millorar-ne la monetització. Aquestes tècniques van des de l'estudi dels indicadors clau del rendiment del videojoc, a mètodes de *machine learning*, passant per anàlisi de dades en temps real o *AB testing*.

L'estadística en el camp dels videojocs gratuïts ha tingut un creixement exponencial en els últims anys, degut a que era un sector en el que, com en molts altres encara actualment, no s'hi havia profunditzat. Les empreses de videojocs F2P han passat de sobreviure sense un anàlisi profund dels seus jocs a tenir departaments dedicats únicament a les *game analytics*, que en el cas de les grans empreses multinacionals poden ser de desenes o centenars de treballadors. Aquests departaments s'encarreguen, en primer lloc, de dissenyar com es volen recollir les dades en el joc. El fet de formar part de la mateixa empresa que el programador que crea el videojoc és un gran avantatge en aquest sentit, ja que són els propis analistes els que decideixen què, quan i com es registra en les seves bases de dades i informen als programadors per què el joc ho fagi d'aquesta manera. També s'ocupen de la gestió i emmagatzemament d'aquestes dades, així com del processament d'aquestes per extreure la informació més valuosa.

El producte o resultat d'aquests departaments són una sèrie d'informes o *reports* i també eines de visualització dels indicadors principals del rendiment del joc per a que les persones que hi tinguin accés puguin saber què està passant-hi en qualsevol moment, en termes generals. Els informes o *reports* que es produeixen són estudis, que es poden fer un sol cop o periòdicament, sobre aspectes concrets del videojoc en els que hi ha més interès, que es presenten als alts càrrecs de l'empresa per tal d'adaptar el joc a aquestes conclusions que s'hagin pogut extreure i així millorar-ne la monetització.

2.1 Indicadors clau del rendiment

Els indicadors clau del rendiment, o KPI, nom que prové de les sigles de *Key Performance Indicator*, són els indicadors que descriuen el funcionament o rendibilitat d'un producte, servei o procés determinat. Existeixen per a qualsevol àmbit, tot i que segons el sector o l'objectiu concret que es vol assolir s'ha d'estudiar unes KPI o altres. Pel que fa a la indústria del videojoc, i concretament pel tipus de videojoc que es tracta en aquest estudi, els *free to play* per a *smartphone*, les KPI principals que veurem es poden classificar en dos grans grups: ús i monetització. Les d'ús corresponen a tot el que té a veure amb el volum d'usuaris que té el joc, usuaris nous, quant juguen, en quina mesura es retenen, etc. Mentre que les de monetització mesuren l'efectivitat a l'hora de generar ingressos que té aquest joc, tant en absolut com en relatiu, pel tal de veure si és rendible dedicar-hi els recursos que són necessaris per a mantenir-lo viu i actualitzat, així com afegir-hi noves funcionalitats³.

³En el sector dels videojocs F2P es poden comprar usuaris mitjançant la publicitat en altres aplicacions, i segons com monetitza el joc en qüestió, pot sortir rendible. Aquest estudi es basa en un joc que no utilitza aquest mètode, si no que viu dels usuaris "orgànics" que genera i en tractar de millorar la seva monetització mitjançant les *game analytics*.

2.1.1 Ús

Les KPI d'ús mesuren una gran quantitat d'aspectes, tot i que es solen centrar en estudiar l'experiència inicial dels usuaris, és a dir, en com es retenen els jugadors en els seus primers passos dins del joc. Aquesta expressió s'utilitza molt ja que unes de les KPI principals i que s'acostuma a tenir molt en compte són les retencions, que veurem a continuació juntament amb la resta.

- **Usuaris actius (DAU, WAU, MAU):**

Els usuaris actius mesuren el volum de persones que juga durant un període determinat. Es mesura en usuaris únics, és a dir, que un usuari que obre el joc dues o més vegades durant el període en qüestió només compta una vegada en els usuaris actius. Així doncs, segons el període en el qual es mesuri el volum d'usuaris s'anomena de forma diferent, segons les sigles en anglès.

El usuari actiu diari s'anomena DAU (*Daily Active Users*), els setmanals WAU (*Weekly Active Users*) i els mensuals MAU (*Monthly Active Users*). Seguint aquest mateix criteri, es poden mesurar els usuaris actius del període que es desitgi.

- **Usuaris nous (NU):**

Els usuaris nous acostumen a anar associats als usuaris actius, és a dir, en el mateix període, tot i que no s'indica en el seu nom com en el cas anterior. Simplement, són els usuaris que han obert el joc per primera vegada durant el període determinat. La proporció del volum d'usuaris actius que són nous és la capacitat que té el joc per captar jugadors, i a partir d'aquí se'ls hi mesura una sèrie de noves KPI agrupant-los segons el període determinat per la seva primera connexió. El grup d'usuaris nous d'un període concret s'anomena la cohort d'aquell període.

- **Retencions (R1, R7, R30):**

Les retencions mesuren la capacitat que té un joc per a fer que els usuaris continuïn jugant dies, setmanes o mesos després d'haver-hi jugat per primer cop. Es calculen sobre les cohorts diàries, i són el percentatge d'usuaris d'una cohort que obra el joc un nombre de dies després de la primera connexió, determinat per la retenció que es vulgui calcular. Així doncs, la retenció a un dia determinat x per a una cohort concreta és:

$$Ret_x = \frac{r_x}{n} \cdot 100$$

On r_x són els usuaris de la cohort que han jugat x dies després de la primera connexió, i n és el nombre d'usuaris que formen la cohort.

Les retencions més comuns a tenir en compte en la indústria són a dia 1, a dia 7 i a dia 30. La retenció a dia 1 (RetD1) és el percentatge d'usuaris d'una cohort que torna l'endemà de la seva primera connexió, a dia 7 (RetD7) és el que torna al cap d'una setmana, i a dia 30 (RetD30) és el que torna al cap de 30 dies. Les retencions es calculen de forma estricta, és a dir que l'usuari només compta que ha retornat si juga el dia concret de la retenció que s'està calculant, tot i que hi jugui dies més tard. Per exemple, un jugador que descobreix el joc un dia determinat, torna a jugar l'endemà i no pot tornar a jugar-hi fins 30 dies després de la seva primera connexió, comptarà per a la RetD1 i la RetD30 de la seva cohort, però no per la RetD7.

- **Mitjana de dies jugats (ADP):**

La mitjana de dies jugats, en anglès *Average Days Played* (ADP), serveix per a mesurar quant enganxa el joc en un moment determinat. En aquest sentit, té el mateix objectiu que les retencions, però en aquest cas es compta el total de dies que han jugat un grup d'usuaris determinat i es divideix pel nombre d'usuaris que conté el grup. Així doncs, l'ADP per a una cohort d'un dia determinat és:

$$ADP_C = \frac{\sum_{i=1}^n \sum_{d=0}^D I_{id}}{n}$$

On D són tots els dies que han passat desde la seva primera connexió, n és el nombre d'usuaris que formen la cohort i I_{id} és una variable binària que val 1 si l'usuari i de la cohort s'ha connectat el dia d o 0 en cas contrari.

Per exemple, si un dia qualsevol entren en el joc 100 usuaris nous i l'endemà en tornen 30, l'ADP d'aquesta cohort en aquest moment serà de $(100+30)/100 = 1.3$. Òbviament, si un grup d'usuaris només hi juga el seu primer dia i ja no tornen, aquest valdrà 1.

D'una manera semblant, es pot calcular l'ADP dels usuaris d'un període concret de la següent forma:

$$ADP_P = \frac{\sum_d DAU_d}{UU}$$

On d representa el nombre de dies del període, DAU_d són els usuaris actius diaris del dia, i UU són els usuaris únics en tot el període.

Així doncs, l'ADP és una mesura més exacta de la quantitat de joc d'un grup d'usuaris determinat, i igual que amb les retencions, es poden segmentar els usuaris i calcular-ne l'ADP per estudiar més profundament el seu comportament. Però l'ADP té una propietat diferent a les retencions, i és que cada dia que passa, el nombre de dies jugats d'una cohort pot augmentar i per tant no és estable. Les retencions a dia 1, dia 7 i dia 30, per exemple, es poden mesurar per a qualsevol cohort que hagi entrat, com a mínim, 1, 7 i 30 dies abans, respectivament. Aquestes retencions un cop calculades no canviaran, encara que els jugadors d'aquesta cohort tornin a jugar més endavant.

2.1.2 Monetització

Les KPI de monetització mesuren la facturació total del joc feta pels usuaris mitjançant les IAP, el volum d'usuaris pagadors i la mitjana de facturació per usuari. Aquesta mitjana es pot calcular de diverses formes, que a continuació veurem, i que serveixen per donar una visió global de com de bé monetitza el joc.

- **Facturació:**

Simplement, la facturació d'un joc és la suma dels ingressos que generen els usuaris. Per a determinar una mica millor de quins usuaris prové aquesta facturació, es pot segmentar en intervals segons diferents criteris, com podria ser l'antiguitat en el joc o el total gastat fins al moment per l'usuari.

- **Usuaris pagadors (PUR):**

El volum d'usuaris pagadors és el total d'usuaris d'un període determinat que ha fet, com a mínim, una compra mitjançant una IAP. S'utilitzen per a calcular el percentatge d'usuaris pagadors respecte el total d'usuaris actius d'un període, que s'anomena *Paying-User Rate*. S'acostuma a calcular diàriament, tot i que es pot calcular també la mitjana ponderada del PUR diàri d'un període determinat utilitzant:

$$PUR_D = \frac{\sum_d payers_d}{\sum_d DAU_d} \cdot 100$$

O bé el PUR global del període:

$$PUR_P = \frac{payers_U}{UU} \cdot 100$$

On $payers_d$ són el nombre de pagadors del dia d , $payers_U$ són el nombre de pagadors únics del període i UU , els usuaris únics.

- **Mitjana de facturació per usuari:**

La mitjana de facturació, en anglès *Average Revenue*, s'acostuma a calcular sobre diferents grups d'usuaris per a donar una visió més ample del comportament. Per un costat, la mitjana de facturació per usuari actiu, normalment diàri o mensual (ARPDau o ARPMAU), per un altre, la mitjana de facturació per usuari pagador d'un període determinat (ARPPU) i, finalment, el que s'anomena *lifetime value* (LTV), que consisteix en una estimació del que es gasta en mitjana cada usuari desde la seva primera connexió fins que abandona el joc. Ho veiem millor a la Taula 1:

KPI	Definició	Fórmula
ARPDau	Mitjana de facturació per DAU	$\frac{F_d}{DAU}$
ARPMAU	Mitjana de facturació per MAU	$\frac{F_m}{MAU}$
ARPPU	Mitjana de facturació per usuari pagador	$\frac{F_p}{payers}$
LTV	Facturació total per usuari únic	$\frac{F_t}{UU}$

Taula 1: Taula resum dels tipus de mitjana de facturació per usuari.

On F_d és la facturació d'un dia determinat, F_m és la facturació d'un mes, F_p és la facturació del període concret pel qual se'n determinin els pagadors únics, F_t és el total de facturació del joc fins al moment i UU és el total d'usuaris que s'ha connectat com a mínim una vegada al joc, els usuaris únics.

2.2 Teorema de descomposició del *free to play*

Aquest teorema descomposa la facturació d'un període qualsevol en varies de les KPI que hem vist anteriorment. És el següent:

$$F_P = UU_P \cdot ADP_P \cdot PUR_d \cdot ARPPU_d$$

On F_P és la facturació total del període, UU_P és el total d'usuaris únics que hi ha hagut en aquest període, ADP_P és la mitjana de dies jugats per aquests, PUR_d és la mitjana ponderada del percentatge de pagadors diari i $ARPPU_d$ és la mitjana ponderada de la mitjana de facturació per usuari pagador diària. Així doncs, veiem com es compleix la fórmula:

$$F_P = UU_P \cdot \frac{\sum_d DAU_d}{UU_P} \cdot \frac{\sum_d payers_d}{\sum_d DAU_d} \cdot \frac{\sum_d F_d}{\sum_d payers_d}$$

Aquesta descomposició ajuda a veure els quatre grans aspectes que determinen els ingressos que genera un videojoc *free to play*. Primer, el volum d'usuaris que hi ha, com més audiència, més probabilitat de captar usuaris pagadors. En segon lloc, la quantitat de joc, mesurat amb l'ADP. Tercer, el percentatge de pagadors, i quart, la mitjana de facturació per usuari pagador, que dependrà del poder adquisitiu d'aquests usuaris i dels preus de les IAP.

De fet, l'ARPPU d'un dia es calcula simplement dividint la facturació entre els DAU, però l'ARPPU mitjà (ponderat segons el volum d'usuaris) d'un període es calcula com:

$$ARPPU_P = \frac{\sum_d F_d}{\sum_d DAU_d}$$

On d són els dies del període, F_d la facturació de cada dia i DAU_d el nombre d'usuaris actius de cada dia. A partir d'aquí, si multipliquem per la suma dels usuaris pagadors diaris al numerador i al denominador, tenim:

$$ARPPU_P = \frac{\sum_d F_d \cdot \sum_d payers_d}{\sum_d DAU_d \cdot \sum_d payers_d}$$

O, escrit d'una altra forma:

$$ARPPU_P = \frac{\sum_d payers_d}{\sum_d DAU_d} \cdot \frac{\sum_d F_d}{\sum_d payers_d}$$

On justament aquests dos factors són les mitjanes ponderades del percentatge de pagadors diari i de l'ARPPU diari en el període en qüestió. Per tant:

$$ARPPU_P = PUR_d \cdot ARPPU_d$$

La conclusió és que l'ARPPU és una mesura de rendiment del joc, però és resultat de vèris factors, alguns incontrolables i imprevisibles per part de l'empresa, en canvi, l'ARPPU, mesurat diàriament en aquest cas, és una mesura que forma part del disseny del videojoc, ja que es pot forçar-lo a variar, per exemple modificant els preus de les IAP. Si es posen ofertes amb preus atractius en el joc, és molt probable que el nombre de pagadors pugi, però també és cert que la mitjana de facturació per usuari pagador baixarà. Així doncs, l'ARPPU d'un joc és una mesura molt valuosa, però s'ha d'anar en compte quan s'estudia.

De la mateixa manera, el producte de l'ARPPDAU descompost per l'ADP dels usuaris d'un període determinat és:

$$PUR_d \cdot ARPPU_d \cdot ADP_P = \frac{\sum_d payers_d}{\sum_d DAU_d} \cdot \frac{\sum_d F_d}{\sum_d payers_d} \cdot \frac{\sum_d DAU_d}{UU_P}$$

Que és el mateix que:

$$PUR_d \cdot ARPPU_d \cdot ADP_P = \frac{\sum_d F_d}{UU_P}$$

Que és justament la definició del *life-time value*. Així que també hem pogut descobrir la descomposició del LTV d'un joc de la següent forma:

$$LTV_P = PUR_d \cdot ARPPU_d \cdot ADP_P$$

2.3 Dades

Les dades que generen els usuaris d'un videojoc gratuït s'acostumen a emmagatzemar a la memòria del dispositiu i a enviar-se per paquets a través de la xarxa quan l'usuari fa unes certes accions. Aquestes dades les reben uns servidors, que poden ser pròpis de l'empresa o contractats d'una altra. Es poden registrar en molts formats diferents, però un dels més eficients i que permet més flexibilitat és el format JSON per línia, que consisteix en guardar cada observació entre dues claus, amb el nom i el valor de cada variable separats per comes. Veiem un exemple a continuació per un usuari qualsevol que fa un inici de sessió.

```
{"id": "12345","event": "inici sessió","data hora": "01-01-2017
00:00:00","plataforma": "Android","pais": "USA","nivell": "10"}
```

El camp "event" permet definir diferents tipus d'esdeveniments (inici de sessió, inici de partida, compra, assoliment d'un objectiu, etc.) i modificar els camps associats a aquest esdeveniment per tal de registrar més informació⁴.

2.3.1 Política de privacitat de dades

Degut a la privacitat de les dades no he pogut accedir a una mostra generada per usuaris d'un videojoc real, ja que en aquest aspecte, hi ha molta competitivitat en el sector i no es volen donar a conèixer xifres ni mètodes utilitzats per a analitzar les dades. En molts casos, ni tant sols els treballadors de la pròpia empresa estan autoritzats a saber el volum d'usuaris que té el joc, o la quantitat d'ingressos que genera.

Aquest fet m'ha portat a crear una simulació amb R en la que els usuaris ficticis es connecten, juguen i compren en el joc, i es registren les seves dades per tal de poder dur a terme l'anàlisi.

⁴El volum de dades que poden arribar a generar els usuaris d'un joc gratuït que funcioni pot ser enorme, de l'ordre de desenes o centenars de gigabytes al dia.

2.3.2 Simulació

La simulació de la base de dades consisteix en una sèrie de funcions, cada una amb un propòsit concret, que es criden en cadena per tal de simular les accions que podria estar fent un usuari en aquest joc fictici.

Així doncs, aquest codi simula l'activitat d'un videojoc F2P que funciona per partides. Això significa que en un moment donat quan es conecta un usuari, que pot ser nou o ja haver jugar anteriorment, aquest pot jugar el nombre de partides que vulgui en el joc, i cada una dura dos minuts. Quan les acaba, se li dona la opció de fer una compra (que es pot entendre com una donació) i, tant si la fa com si no, se li actualitza la informació del seu perfil amb l'activitat que ha generat. A més, es guarden dos variables referents al país de procedència i la plataforma que utilitza el dispositiu d'aquest usuari fictici.

Les funcions que s'utilitzen en la simulació es defineixen a continuació:

- **simulation**: Aquesta funció és la que es crida per a fer la simulació, i simplement, executa la funció `one.day` tantes vegades com dies es vulguin simular, amb el paràmetre `ndays`. A part, calcula el temps d'execució per a fer la simulació i l'escriu per pantalla, juntament amb altres característiques de la simulació que es poden modificar, com la duració de cada dia fictici, el nombre de dies, els països d'on poden provenir els usuaris i la plataforma del dispositiu.
- **one.day**: Genera un dia de simulació. Els paràmetres `day` i `ndays` permeten a la funció `is.weekend` determinar si el dia és cap de setmana o no, ja que les probabilitats d'entrada de jugadors canvien. A continuació executa un bucle `while` que simula el dia i el temps mentre va passant. A partir d'aquí, es simula l'entrada, les partides i les compres dels usuaris que van entrant, segons unes certes probabilitats que es defineixen més endavant, amb les funcions `user`, `behavior`, `play` i `iap`.
- **is.weekend**: Estableix la probabilitat d'entrada d'un usuari en un moment donat, i també la de que sigui un usuari nou, segons si el dia en qüestió cau en cap de setmana o no, suposant que la simulació comença en el dia 1 un dilluns.
- **user**: Retorna l'identificador d'un usuari que pot ser nou o haver jugat anteriorment. Si és nou, s'afegeix a la base de dades on es registra la informació de tots els usuaris que han entrat (`user.info`), utilitzant la funció `profile` amb el paràmetre `type="new"`. Si l'usuari és antic, s'escull un de la base de dades dels usuaris i s'actualitza la informació del seu perfil cridant a la mateixa funció `profile` però amb el paràmetre `type="existing"`.
- **profile**: Per als usuaris existents, actualitza el seu total de sessions, el total de dies jugats i el compta com a DAU, si és la primera sessió del dia. Per als usuaris nous, genera el seu perfil seleccionant a l'atzar un país i una plataforma, i també el seu tipus de comportament amb la funció `behavior` amb el paràmetre `type="new"`.
- **behavior**: Pel `type="new"` defineix el tipus de comportament de l'usuari segons unes certes probabilitats, que pot ser *spontaneous*, *casual*, *com-*

pulsive o *whale*. Pel `type="probs"` retorna les probabilitats de repetir partida i de fer una compra segons el tipus de comportament de l'usuari.

- **play:** Decideix, segons el paràmetre de repetir partida, el nombre de partides que fa l'usuari amb la funció `rtpois`. En aquest joc, l'usuari puja un nivell cada partida que fa, així que aquesta funció també s'encarrega d'actualitzar el nivell en la base de dades dels usuaris.
- **rtpois:** Genera un nombre aleatori segons una distribució de poisson zero-truncada, ja que l'usuari com a mínim juga una partida quan es connecta, mitjançant el mètode de l'invers. Aquest mètode consisteix en generar una mostra aleatòria d'una variable X amb distribució de probabilitat $F_X(x)$ a partir d'una variable U amb distribució uniforme $U(0, 1)$, simulant una realització de $F_X^{-1}(u)$.
- **iap:** La funció per a fer les compres primer de tot mira si l'usuari està registrat com a gratuït o pagador, amb el camp `UserType` de la base de dades `user.info`. Si és gratuït, el canvia a pagador. A continuació, decideix aleatòriament quina compra fa l'usuari, que hi ha cinc preus diferents, i li suma aquesta quantitat tant al seu total gastat a la base de dades dels usuaris, com en la facturació del dia en la base de dades diària. Si és la seva primera compra del dia, també el compta com a pagador únic en la base de dades diària.

El resultat de la simulació amb els paràmetres utilitzats és el següent:

```
Time difference of 1.937173 hours
Parameters used:
Day duration (in seconds): 86400
Number of days: 365
Possible countries: XX YY ZZ
Possible platforms: iOS Android
```

Consisteix en dues bases de dades: una amb la informació del joc diàriament, anomenada `daily`, i l'altra amb la informació de tots els usuaris ficticis que hi han jugat, `user.info`. La primera base de dades conté, per cada dia de simulació, els DAU, els usuaris nous, la facturació, l'ARPDau, el nombre d'usuaris pagadors, l'ARPPU i el nombre d'usuaris que han retornat 1, 7 i 30 dies després de la seva primera connexió. Veiem un exemple a la Taula 2:

DAU	NU	Revenue	ARPDau	Payers	ARPPU	R1	R7	R30
113	113	406.65	3.599	28	14.523	0	0	0
239	132	458.58	1.919	35	13.102	107	0	0
291	101	669.54	2.301	42	15.941	103	0	0
334	117	461.69	1.382	27	17.100	67	0	0
404	132	530.61	1.313	33	16.079	67	0	0

Taula 2: Mostra de la base de dades diària.

La segona base de dades conté el identificador de l'usuari, el país des del que juga, la plataforma del dispositiu, si és pagador o gratuït, el dia en el que va jugar per primer cop, la variable que descriu el comportament de l'usuari, el nivell assolit, el total de dies que ha jugat, el total de sessions i el total que s'ha gastat mitjançant les IAP. Veiem l'exemple a la Taula 3:

id	país	platform	type	first	behavior	lvl	DP	TS	TRev
4305	YY	iOS	freeloader	28	compulsive	6	4	4	0
776	ZZ	iOS	freeloader	6	compulsive	12	7	8	0
37513	YY	iOS	freeloader	246	casual	2	1	1	0
7975	ZZ	Android	freeloader	53	casual	5	4	4	0
26280	XX	Android	payer	174	compulsive	11	3	5	1.99

Taula 3: *Mostra de la base de dades dels usuaris.*

El fet de generar la base de dades `user.info` en l'entorn global i la base de dades `daily` en un entorn temporal dins de la funció, és per tal de que es pugui generar un període amb unes determinades probabilitats (d'entrada, de compra, etc.) i després simular un altre període amb probabilitats diferents però amb la mateixa base d'usuaris, a part dels nous que vagin entrant. D'aquesta manera es podria estudiar com canvia el comportament dels usuaris envers un canvi dins del joc, tot i que no s'ha pogut dur a terme en aquest estudi, per a profunditzar en l'anàlisi de com afecten aquestes modificacions o nous continguts.

3 Anàlisi

Un cop s'ha simulat les dues bases de dades, la part d'anàlisi comença per estudiar les KPI sobre el rendiment del videojoc en global, i una vegada s'ha fet, es pot profunditzar en l'estudi del perfil dels usuaris per a entendre millor per què actuen com actuen. Aquesta secció consta de tres apartats, en el primer es descriuen totes les mètriques relacionades amb el funcionament del joc i la simulació en global (volum d'usuaris, les seves retencions, la facturació que ha generat, etc.), el segon es centra en les característiques d'ús del joc que tenen els usuaris (d'on provenen, com es comporten, la mitjana de dies jugats, etc.), i el tercer estudia en profunditat la monetització del joc segons els diferents perfils d'usuari (percentatge de pagadors, mitjana de facturació, etc.)

3.1 Estudi del rendiment d'un *free to play*

3.1.1 Mètriques d'ús

Els indicadors clau d'ús informen del volum d'usuaris del joc i de com es comporten envers aquest. A continuació s'analitzaran aquests indicadors pels usuaris que hi ha hagut en la simulació del joc, començant pel volum d'usuaris que ha tingut.

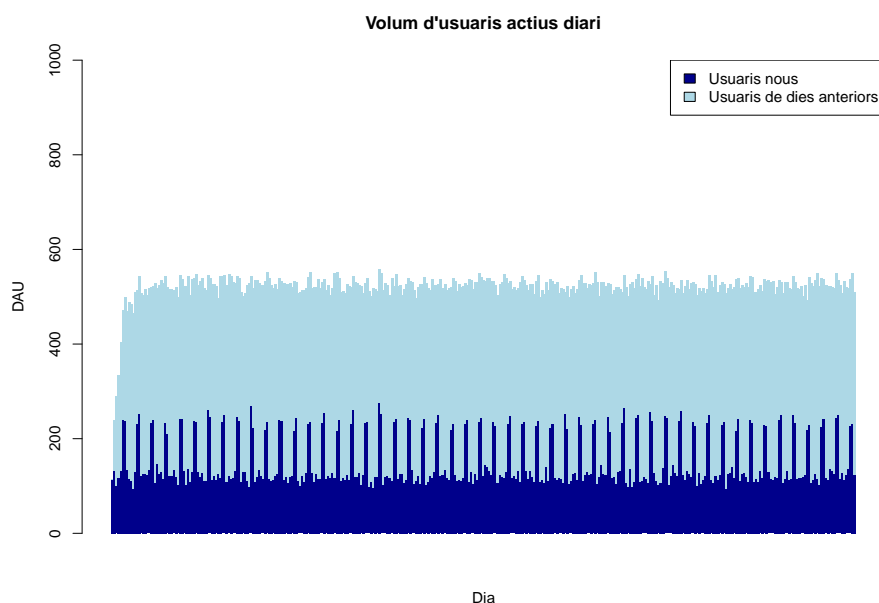


Figura 5: *Volum d'usuaris actius diari amb usuaris nous.*

Com es pot observar a la Figura 5, aquest joc té al voltant de 500 DAU, amb un nombre d'usuaris nous diaris que varia segons el dia de la setmana. Els caps de setmana, degut a que la majoria de la població té més temps lliure, el volum d'usuaris nous puja per sobre els 200 cada dia, mentre que els dies entre setmana es queden entre 100 i 150. Passem a inspeccionar visualment les

retencions a dia 1, 7 i 30 de les cohorts diàries d'aquest joc. En la indústria, es considera que un videojoc *free to play* reté prou bé els usuaris nous si té la retenció a dia 1 del 40%, la retenció a dia 7 del 20% i la retenció a dia 30 del 10%. Normalment, la retenció que més s'allunya d'aquests estàndards és la a dia 30, ja que és molt difícil fer millores en l'experiència inicial dels usuaris que siguin efectives durant tants dies.

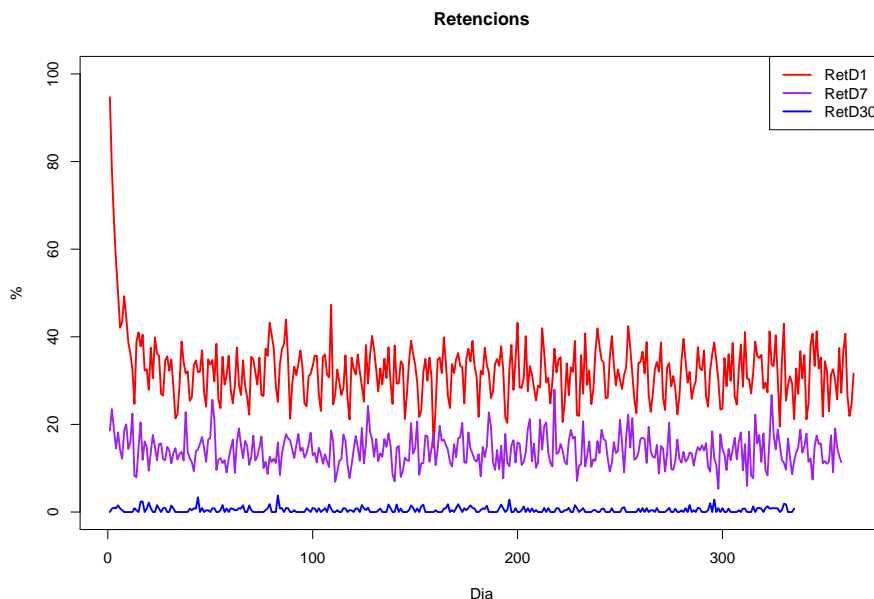


Figura 6: Retenció a dia 1, 7 i 30.

Com és lògic, la retenció a dia 1 arriba fins el dia anterior a la finalització de la simulació, la retenció a dia 7 fins una setmana abans, i la retenció a dia 30 fins un mes abans, ja que no han passat suficients dies per saber si els usuaris de les últimes cohorts han tornat el nombre de dies després. Les retencions d'aquest hipotètic joc són una mica inferiors al que es recomana, tot i que no són gens dolentes, en especial a curt termini. La retenció a dia 1 global del període que s'ha simulat, calculada amb la mitjana de les retencions a dia 1 diàries ponderada segons la mida de la cohort, és 31.73%. La retenció a dia 7 global és 13.51%, mentre que la retenció a dia 30 global és 0.5%.

3.1.2 Mètriques de monetització

La monetització d'un videojoc *free to play* està determinada per varis aspectes, com hem vist anteriorment, però el producte d'aquests, la facturació absoluta, és el que determina si aquest joc és rendible o no. Al cap i a la fi, els ingressos que genera són el més important com en qualsevol negoci, tot i que per tal d'augmentar-los es poden seguir diferents estratègies. Veiem la facturació diària que ha generat el joc fictici que hem simulat a la Figura 7:

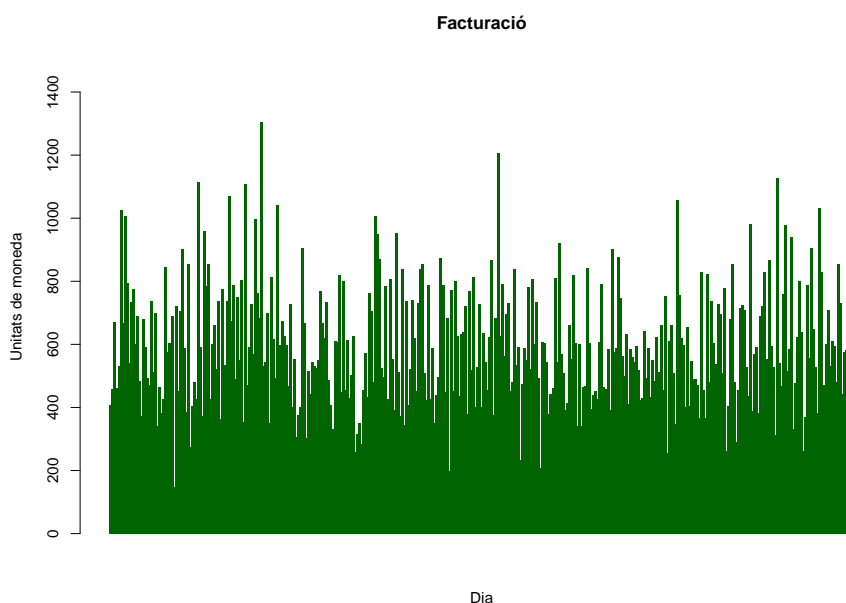


Figura 7: *Facturació diària mitjançant IAP's.*

La facturació total que ha generat aquest joc en el període de 365 dies ha sigut de 217574 unitats de la moneda que s'utilitzi, que acostuma a ser en *dollars* (\$). En cada país les IAP es fan a través del seu propi servidor del mercat d'aplicacions, i per tant els preus que es mostren en el joc i que reben les empreses són en la moneda local. Com que en aquest tipus de jocs hi pot jugar gent de tot el món en qualsevol moment, a l'hora de registrar les dades, un factor molt important és guardar el codi de la moneda en el que es registra la quantitat gastada per l'usuari, no és el mateix una compra de \$100 que una de 100¥ (yens japonesos). En general, les empreses mesuren els ingressos dels seus jocs en *dollars*, tot i tenir les seves seves en països on el *dollar* no és la moneda oficial. Continuant amb les KPI de monetització, veiem el volum d'usuaris pagadors diari que hi ha hagut en la simulació en la Figura 8.

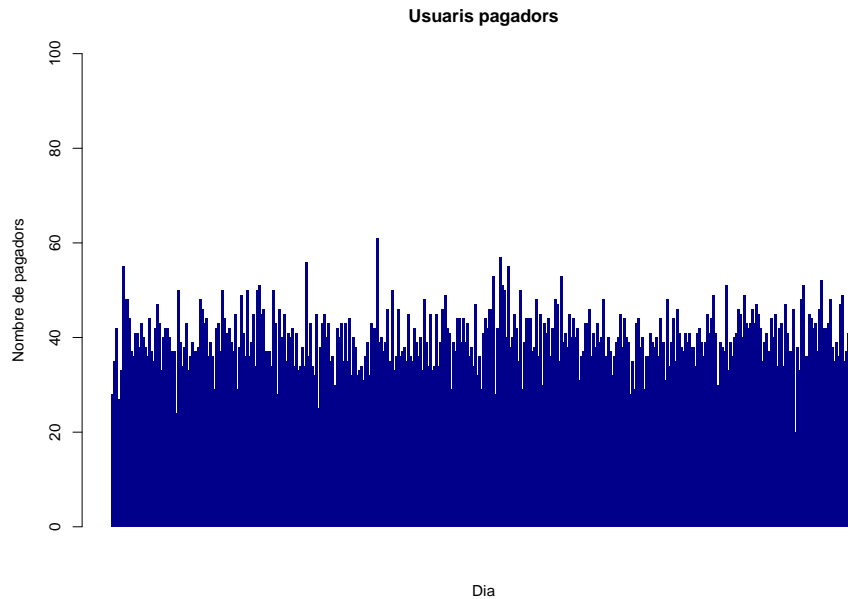


Figura 8: *Nombre d'usuaris pagadors diari.*

Veiem com el nombre d'usuaris pagadors que hi ha hagut diàriament en la simulació del joc ha estat al voltant de 40.

Passem a veure la mitjana de facturació per usuari, calculada de dues formes diferents per entendre millor com s'ha generat la facturació que hem vist provinent dels usuaris pagadors anteriors, per usuari actiu diari (ARPPU) i per usuari pagador diari (ARPPU). Aquestes dues mitjanes acostumen a ser molt diferents, l'ARPPU global de la simulació es situa en 1.144, mentre que l'ARPPU diari global és de 14.891. Així com és relativament senzill trobar una frase per a descriure la mitjana per usuari pagador diari ("els usuaris pagadors d'aquest joc es gasten en mitjana 15 *dollars* cada dia"), no hi ha una manera intuïtiva per a descriure l'ARPPU. Pot resultar enganyós també, ja que incorpora la informació del volum d'usuaris actius, d'usuaris pagadors i la quantitat gastada per aquests, i es pot malinterpretar fàcilment atribuïnt erròniament el seu comportament a un fet determinat. Veiem aquestes dues mitjanes diàries de la simulació que hem fet.

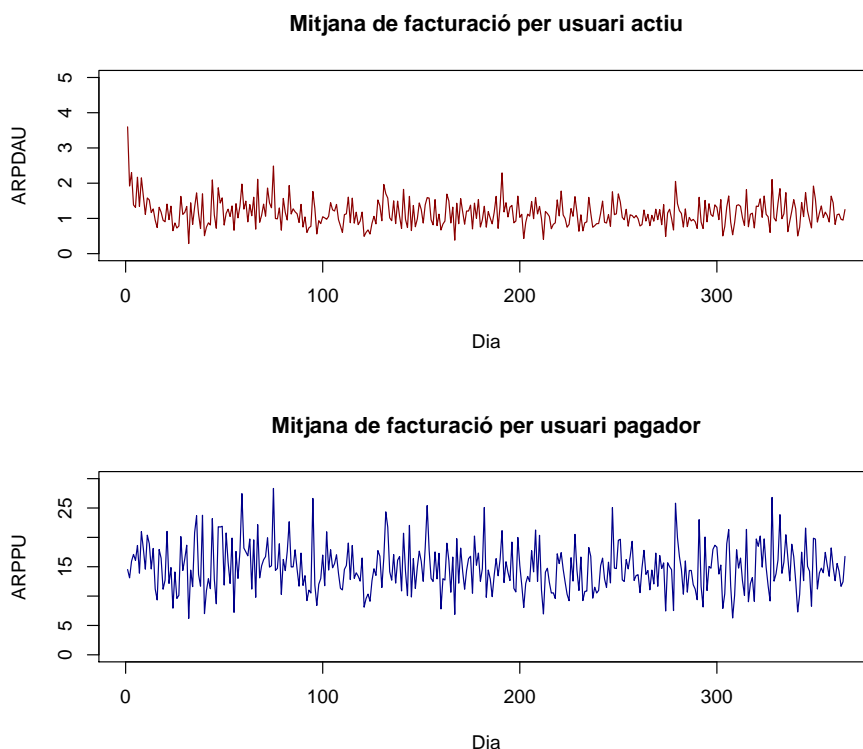


Figura 9: *Mitjanes de facturació per usuari diàries.*

A la Figura 9 es veu com s'estabilitzen entorn els valors que hem calculat anteriorment, i lògicament, la variabilitat de la mitjana de facturació diària és molt més alta en el cas de l'ARPPU, ja que el nombre d'usuaris pagadors varia en major proporció que els DAU.

3.2 Estudi del perfil d'ús dels usuaris

El primer que s'ha de fer per a endinsar-se en l'estudi i la comprensió del comportament dels usuaris és fer un anàlisi descriptiu de les variables que s'han registrat. Comencem veient com es classifiquen els usuaris segons el país de procedència i la plataforma que utilitza el seu dispositiu.

	iOS	Android	Total
XX	9204	9256	18460
YY	9454	9304	18758
ZZ	9344	9127	18471
Total	28002	27687	55689

Taula 4: *Nombre d'usuaris segons país de procedència i plataforma.*

Com es pot veure a la Taula 4, el nombre d'usuaris que provenen de cada un dels tres països ficticis que incorpora la simulació està uniformement distribuït, el mateix que si ho mirem segons la plataforma del seu dispositiu i també si creuem ambdues variables. Aproximadament, hi ha el mateix nombre d'usuaris de cada país i plataforma.

A continuació passem a veure els quatre tipus de comportaments que poden tenir els usuaris. Els *spontaneous* són els usuaris que no formen part del *target* del joc, o simplement que no els agrada, ja que són usuaris que la majoria de vegades es connecten, proven el joc fent una partida i ja no tornen més, i gairebé mai fan cap compra. Els *casual* són usuaris que sí que són del públic objectiu del joc, tot i que juguen ocasionalment (per exemple, els cap de setmana) i durant poca estona, però continuen jugant amb el pas del temps i, de tant en tant, algun fa una compra. Els *compulsive* serien els usuaris més *hardcores*, els que descobreixen el joc i hi juguen molt fins que se'n cansen. Aquests usuaris poden generar quantitats de facturació importants, ja que són els que dediquen més hores en el joc. Finalment, els *whales* és com s'anomena als jugadors que, sense estar molt enganxats ni jugar desde fa molt de temps, es gasten quantitats enormes de diners, fins a punt de comprar-se tot el que hi ha disponible en un joc. Veiem com es distribueixen els usuaris d'aquesta simulació en funció d'aquesta classificació.

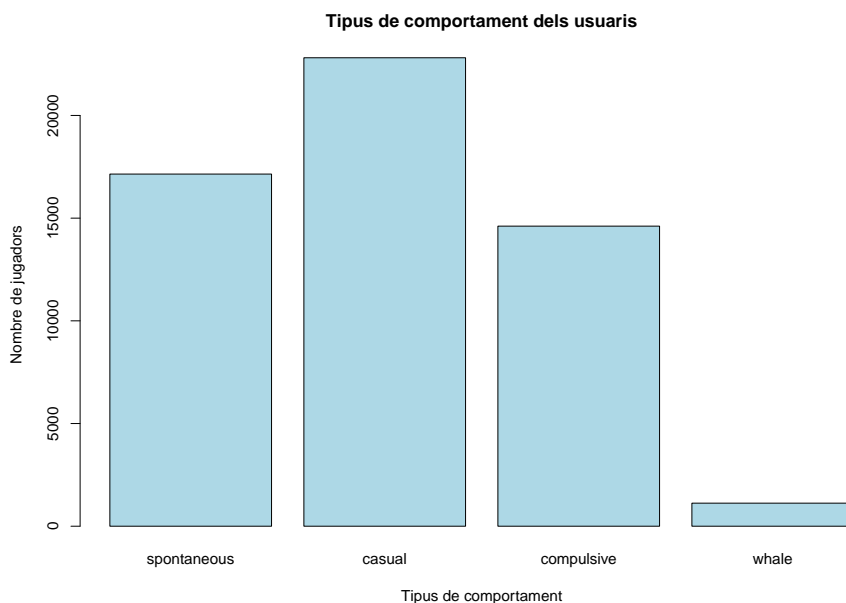


Figura 10: Nombre de usuaris segons el tipus de comportament.

A la Figura 10 s'hi veu que el grup amb més usuaris és el de jugadors *casual*, seguit dels *spontaneous* i els *compulsive*. Com és lògic, el nombre d'usuaris *whales* és prou baix, però cada un d'aquests pot ser més rendible que 10 jugadors dels altres tipus de comportament. Procedim a estudiar aquests quatre tipus de comportament en funció del país i la plataforma que utilitzen els usuaris, per

veure si podem considerar que les poblacions de cada país o de cada plataforma tenen comportaments diferents.

Tipus	Plataformes		Països		
	iOS	Android	XX	YY	ZZ
<i>spontaneous</i>	8561	8585	5623	5775	5748
<i>casual</i>	11528	11281	7634	7706	7469
<i>compulsive</i>	7332	7279	4834	4885	4892
<i>whale</i>	581	542	369	392	362

Taula 5: Nombre d'usuaris segons el tipus de comportament, en cada país i plataforma de la simulació.

L'única conclusió que podem extreure de l'estudi de la Taula 5 és que ni el país ni la plataforma del dispositiu dels usuaris influeixen significativament en la distribució dels seus tipus de comportament, ja que els valors de cada un són molt propers entre països i entre plataformes.

Pel que fa al nivell dels usuaris, que és el mateix que el nombre total de partides que ha jugat cada un, es distribueix de la següent forma:

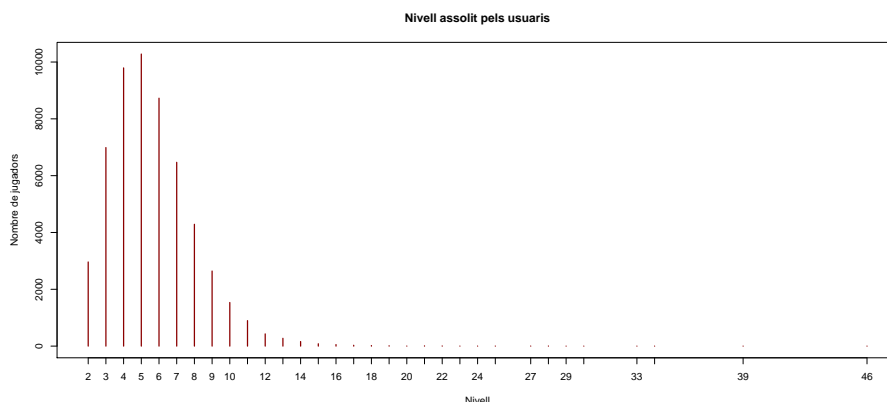


Figura 11: Nombre de usuaris segons el nivell assolit en tot el seu historial de joc.

Com hem vist a l'explicació de la simulació, cada usuari que es connecta comença sent nivell 1 i juga, com a mínim, una partida en la seva primera sessió. Per tant, és lògic que el nivell mínim que hi ha sigut 2, pels usuaris que només hagin jugat una partida i no hagin tornat més. Els nivells on arriben la majoria dels usuaris són el 3, 4, 5 i 6, i a partir d'aquí, el nombre d'usuaris per nivell va decreixent de forma exponencial fins a arribar al punt que hi ha molt pocs usuaris per sobre del nivell 30. Tot i així, s'hi veuen usuaris que han assolit el nivell 33, 34, 39 i 46 respectivament.

Passem a estudiar la relació entre el nivell assolit pels usuaris i les tres variables que hem descrit anteriorment. Per a fer-ho, utilitzem un tipus diferent de gràfic, en el qual es representa en l'eix horitzontal el nivell assolit, i en l'eix vertical els possibles valors de la variable categòrica amb la que volem creuar-lo.

Veiem l'exemple pel cas del país des del que juga l'usuari.

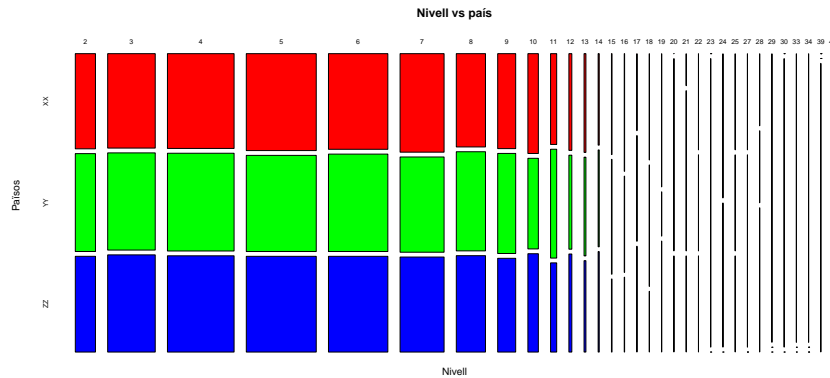


Figura 12: *Distribució del nivell assolit pels usuaris segons el país de procedència.*

L'amplada de cada columna representa el volum d'usuaris que han arribat fins aquell nivell, mentre que la divisió en colors representa la proporció dels usuaris d'aquell nivell que pertanyen a cada valor de la variable categòrica, en aquest cas a cada país. En aquesta gràfica no s'hi aprecia cap canvi en la distribució del país d'on prové l'usuari segons el nivell assolit per aquest. Seguim el mateix procediment per a estudiar la relació amb la plataforma del dispositiu i amb el tipus de comportament de l'usuari.

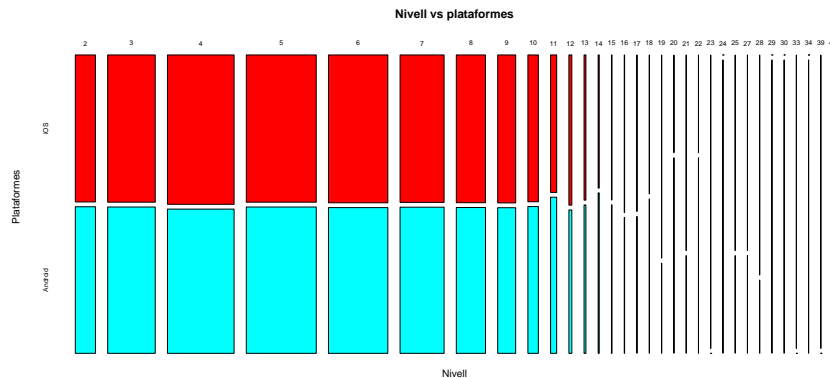


Figura 13: *Distribució del nivell assolit pels usuaris segons la plataforma dels dispositiu.*

Igual que en el cas anterior, en la Figura 13 tampoc s'hi veu cap relació evident entre el nivell al que arriben els usuaris i la plataforma que utilitza el seu dispositiu. Fem el mateix per la variable que descriu el comportament de l'usuari.

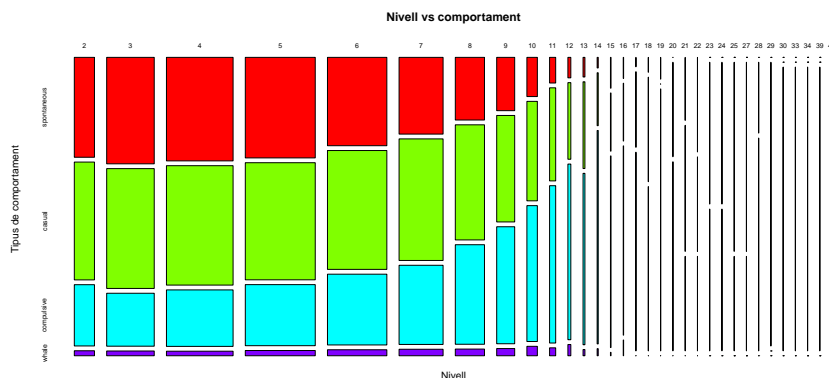


Figura 14: *Distribució del nivell assolit pels usuaris segons seu tipus de comportament.*

En aquest cas veiem per primera vegada un aspecte que està relacionat amb el nivell assolit, i és el tipus de comportament. Com hem vist anteriorment en la definició de cada un dels tipus de comportament dels usuaris, aquests tendeixen a jugar més o menys partides segons com siguin, i per tant, els usuaris que juguen més arribaran a un nivell més alt. Mirant la gràfica, veiem que en els nivells del 3 al 6 s'hi troba la majoria dels usuaris, com ja sabem, però també que una gran part d'aquests són *spontaneous* o *casual*. A mesura que ens fixem en nivells més alts, aquesta distribució canvia i el tipus de comportament que predomina és el *compulsive*, ja que són els que tendeixen a jugar més. La proporció de *whales* sembla que és constant al llarg dels nivells.

Canviem d'aspecte per estudiar la mitjana de dies jugats pels usuaris. En global, l'ADP d'aquest joc és de 3.41 dies jugats per usuari, tot i que s'entén molt millor el fenomen estudiant-lo per cohorts o segmentat segons altres criteris. El veiem per cohorts diàries en la Figura 15.

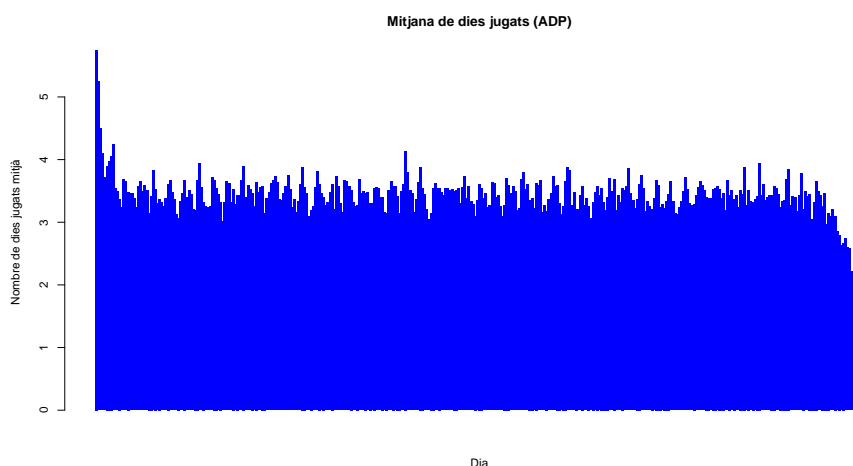


Figura 15: *Mitjana de dies jugats per cohort.*

Veiem que l'ADP pels usuaris que van entrar durant els primers dies del joc és el més alt, després s'estabilitza al voltant de 3.5, i finalment baixa a mesura que el dia d'entrada de la cohort és més recent. Aquest comportament és totalment lògic, ja que els usuaris dels primers dies han disposat de més temps per a jugar. També ho és el comportament pseudo-sinusoidal que té durant el període en el que està estable, i correspon als caps de setmana. Al captar més jugadors nous els caps de setmana fa que també entrin usuaris que no formen part del *target* del joc, i per tant l'abandonen abans i fan baixar la mitjana de dies jugats.

3.3 Estudi del perfil de monetització dels usuaris

Per estudiar la monetització dels usuaris que han jugat en el nostre joc fictici en profunditat començarem amb el percentatge de pagadors, que hem pogut veure anteriorment amb el volum d'usuaris pagadors que hi ha hagut. Per a fer-ho, utilitzarem la variable `Usertype` de la base de dades dels usuaris per a calcular el percentatge de pagadors global, i segons varis factors.

Freeloaders	Payers	Total
45102	10587	55689

Taula 6: Nombre d'usuaris segons si són gratuïts o pagadors.

El que veiem en la Taula 6 bàsicament és el nombre d'usuaris de cada tipus, i amb aquestes dades podem extreure el percentatge de pagadors que han fet una compra com a mínim, en tot el període de la simulació, el que anteriorment hem anomenat PUR global que és de 19.01%. Es pot calcular també el PUR diari mitjà amb la fórmula que hem vist i és de 7.68%. A continuació veurem el percentatge de pagadors segons varies característiques del seu perfil, començant per el país i la plataforma.

Plataformes		Països		
iOS	Android	XX	YY	ZZ
18.81	19.22	19.18	18.99	18.87

Taula 7: Percentatge d'usuaris pagadors en cada país i plataforma de la simulació.

Veiem com no hi ha diferències significatives entre el percentatge d'usuaris pagadors segons el país des d'on juga o la plataforma del seu dispositiu. En canvi, si ho mirem segons el tipus de comportament dels usuaris, el percentatge de pagadors és el següent:

<i>spontaneous</i>	<i>casual</i>	<i>compulsive</i>	<i>whale</i>
3.95	17.7	32.51	100

Taula 8: Percentatge de pagadors global segons el tipus de comportament.

Aquest percentatge és molt petit pels usuaris espontanis, i augmenta en funció que el seu tipus de comportament es torna més radical, fins que els

whales ja són tots grans gastadors. Veiem-lo també en funció del nivell assolit pels usuaris en la Figura 16.

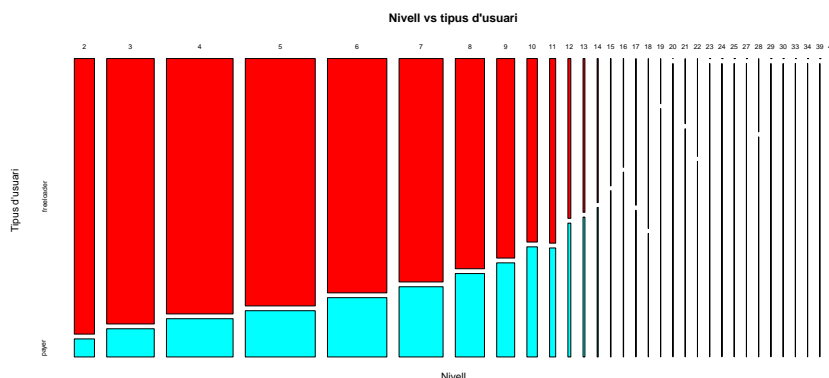


Figura 16: Distribució del nivell assolit pels usuaris segons si són gratuïts o pagadors.

Mirant el percentatge de pagadors segons el nivell, veiem com a mesura que l'usuari puja de nivell, hi ha una major proporció d'aquests que són pagadors. És un fet lògic també, ja que els usuaris que han assolit un nivell més alt és perquè han jugat més, i per tant han tingut més oportunitats de fer IAP's.

Si es vol estudiar la mitjana de facturació segons algun d'aquests factors, es calcula la mitjana de facturació per usuari (ARPU) per a cada nivell de la variable, el veiem a continuació segmentat per país (Taula 9), plataforma del dispositiu (Taula 10) i comportament (Taula 11).

	XX	YY	ZZ
ARPU	3.875	3.816	4.032

Taula 9: Mitjana de facturació per usuari segons el país de procedència.

Veiem com els usuaris del país ZZ tenen una mitjana de facturació lleugerament més elevada que els dels altres països. Veiem-ho per plataformes.

	iOS	Android
ARPU	3.922	3.892

Taula 10: Mitjana de facturació per usuari segons la plataforma del seu dispositiu.

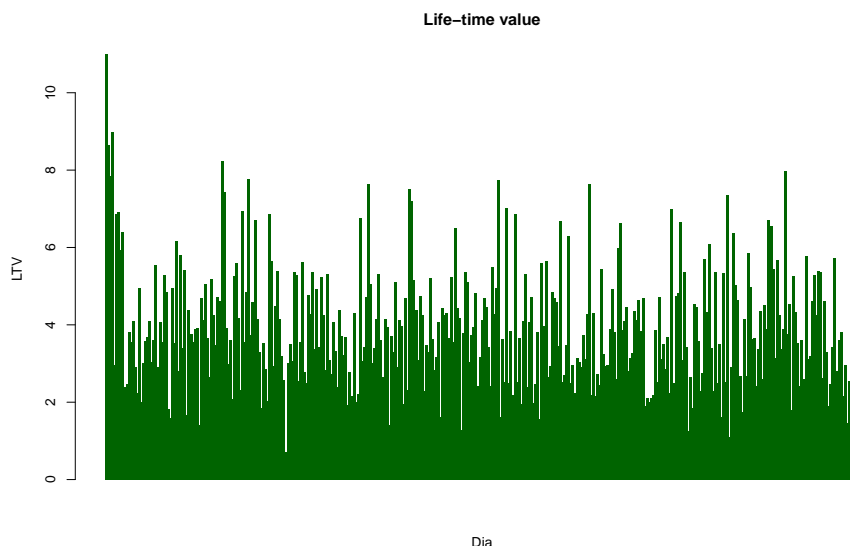
És molt propera entre les dues plataformes disponibles. En el cas de la mitjana de facturació segons el tipus de comportament de l'usuari sí que podem apreciar una evolució.

En la Taula 11 s'hi veu com la mitjana de facturació pels usuaris espontanis és molt baixa, menor de 1, lleugerament inferior a 3 pels usuaris casuals, 5.5 pels que juguen compulsivament i més de 50 pels usuaris *whales*.

	<i>spontaneous</i>	<i>casual</i>	<i>compulsive</i>	<i>whale</i>
ARPU	0.553	2.817	5.506	56.452

Taula 11: Mitjana de facturació per usuari segons el tipus de comportament.

Finalment, s'ha calculat el *life-time value* d'aquest joc, dividint el total de facturació que ha generat en la seva història entre el total d'usuaris únics que hi ha jugat, i és de 3.907 unitats. Igual que hem fet anteriorment amb la mitjana de dies jugats, a continuació calculem el LTV de les cohorts diàries d'aquest joc. Això es fa sumant el total que ha ingressat el joc dels usuaris nous de cada un dels dies de la simulació. Per exemple, si un usuari que entra el primer dia de simulació fa una compra i una setmana després en fa una altra, la facturació corresponent a aquestes dues compres es sumarà a la facturació de la cohort del primer dia. Així doncs, dividint la facturació total de cada cohort entre el nombre d'usuaris que conté obtenim el *life-time value* per cohort diària.

Figura 17: *Life-time value* per cohorts diàries.

En la Figura 17 veiem com els usuaris que van entrar per primera vegada els primers dies de la simulació tenen un LTV més elevat, mentre que els més recents el tenen molt més baix, com ja es podia esperar, sobretot degut al temps que porten jugant. Tot i així, també s'hi veuen dies concrets prou avançats en que el LTV és superior a 6, fet que es deu a que probablement algun usuari d'aquella cohort ha fet una compra de les que tenen un preu més elevat, i per tant la mitjana de la seva cohort s'ha vist altament afectada.

Així conclou la part d'anàlisi de les base de dades que ha generat la simulació, intentant emular el que es podria registrar i analitzar sobre un videojoc *free to play* que estigui actiu.

4 Conclusions

En aquest estudi sobre l'anàlisi de dades en el món dels videojocs gratuïts per a dispositius intel·ligents, conegut com *game analytics*, s'ha pogut veure una introducció al sector dels videojocs, amb els fets històrics rellevants que han afectat en l'evolució d'aquesta indústria, així com tot el procés que s'ha donat per arribar a la aparició de jocs *free to play*. També s'han definit els indicadors clau del rendiment (KPI) que s'estudien en el seguiment d'un videojoc, així com possibles factors del perfil dels usuaris que afecten a aquests indicadors.

Hem vist també com analitzar la monetització d'un joc gratuït, mesurant indicadors relatius com el PUR, l'ARPDau o l'ARPPU, i veient com poden afectar els canvis en el joc en aquests indicadors. Un aspecte molt important sobre la monetització d'un videojoc és la descomposició de la facturació que hem vist en la Secció 2.2, el que s'anomena teorema de descomposició del *free to play*. D'acord amb aquesta descomposició, per millorar la monetització d'un joc gratuït podem centrar-nos en qualsevol dels quatre aspectes que la formen: volum d'usuaris, mitjana de dies jugats, percentatge de pagadors i mitjana de facturació per usuari pagador. Si es vol simplement augmentar el volum d'usuaris del joc, es pot utilitzar la publicitat, tot i que si s'utilitza la compra d'usuaris a altres aplicacions o jocs les KPI que s'ha d'estudiar són unes altres (com podria ser el cost mitjà per adquisició i comparar-lo amb el LTV). La mitjana de dies jugats conté la informació sobre l'entreteniment que proporciona el joc, i quan tarden els usuaris a abandonar-lo. L'única forma de millorar-la és afegint continguts en el *gameplay* del joc. Finalment, el percentatge de pagadors i la mitjana de facturació per usuari pagador estan molt relacionats, com hem pogut veure en l'apartat de l'estudi de l'ARPDau. La millor forma de millorar la monetització en aquests aspectes seria buscar la combinació de preus de les IAP que hi ha en el joc, que fa òptima la relació entre PUR i ARPPU.

També hem vist en la Secció 2.3.1 que les dades sobre videojocs reals tenen una confidencialitat màxima, i que és molt difícil accedir a una mostra per a realitzar un estudi de la monetització d'un joc, sobretot amb finalitats acadèmiques. Com que l'estadística en aquest sector encara s'està desenvolupant, no hi ha uns estàndards de com monetitzar un videojoc o de com analitzar les dades per treure'n la màxima rendibilitat, i l'empresa que dona amb la clau per a fer-ho ho manté en secret. D'aquesta manera, hem pogut veure com es pot fer una simulació d'un videojoc en R, intentant ser el més fidel possible a com es comporten els usuaris d'un joc en realitat.

En l'anàlisi de les dades (Seccions 3.1, 3.2 i 3.3), hem descobert que hi ha factors que influeixen en el comportament dels usuaris envers el joc i les compres, i així com, per exemple, els usuaris d'aquest videojoc fictici es comporten igual segons el país des d'on juguen (fet que no té per què donar-se a la realitat), l'antiguitat en el joc o el tipus d'usuari que són afecta a la facturació que genera aquest joc. Aquest anàlisi pot donar lloc a decisions com podrien ser: que l'empresa es dediqui exclusivament a millorar la jugabilitat del joc, ja que no té prou contingut, posar publicitat per a augmentar la massa d'usuaris, o modificar els preus de les IAP que s'ofereixen en el joc.

Dur a terme aquest anàlisi i aplicar les conclusions que se'n extreguin pot provocar una millora en la monetització i rendibilitat d'un videojoc gratuït, i per tant, pot ser un factor a tenir molt en compte per a les empreses que es dediquen a crear i mantenir aquest tipus de videojocs.

5 Bibliografia

1. King, Lucien. "*Game on: the history and culture of videogames*", Laurence King, 2002.
2. M. Ross, Sheldon. "*Simulation*" chapter 4, Elsevier, 5th edition, 2013.
3. Magy Seif El-Nasr, Anders Drachen and Alessandro Canossa. "*Game Analytics: Maximizing the value of player data*" Springer, 2013.

6 Annexes

```
##### FUNCIONS #####
##### PRINCIPAL
simulation <- function(tmax,ndays){
  time.begin <- Sys.time()
  results <- NULL
  for(i in 1:ndays){
    results[[i]] <- one.day(i,ndays)
  }
  names(results) <- paste("Day",1:ndays,sep=" ")
  time.end <- Sys.time()

  print(time.end-time.begin)
  cat("Parameters used:
  Day duration (in seconds):",tmax,"
  Number of days:",ndays,"
  Possible countries:",countries,"
  Possible platforms:",platforms,"\n")

  return(results)
}

one.day <- function(day,ndays){
  probs <- is.weekend(day,ndays)
  env$day.info <- data.frame(DAU=numeric(1),NU=numeric(1),
    Revenue=numeric(1),ARPPDAU=numeric(1),Payers=numeric(1),
    ARPPDAU=numeric(1),R1=numeric(1),R7=numeric(1),R30=numeric(1))
  env$uids <- numeric(0)
  env$payers <- numeric(0)

  t <- 1
  while(t < tmax){
    p <- runif(1)
    if(p < probs[1]){
      uid <- user(user.info,day,probs[2])
      probs.class <- behavior(type="probs",user.info[uid,"Behavior"])
      # MATCHES
      ngames <- play(uid,probs.class[1])
      t <- t + ngames*gametime
      # IAPs
      p.iap <- runif(1)
      if(p.iap < probs.class[2]){
        iap(uid)
        t <- t + iaptime
      }
    }
  }
  t <- t + 1
}
```

```

}
env$day.info$ARPDau <- round(env$day.info$Revenue/
  env$day.info$DAU,3)
env$day.info$ARPPDAU <- round(env$day.info$Revenue/
  env$day.info$Payers,3)
return(env$day.info)
}

##### FUNCIO IS WEEKEND
is.weekend <- function(day,ndays){
  weekends <- sort(c(seq(6,ndays,by=7),seq(7,ndays,by=7)))
  if(day %in% weekends == TRUE) return(c(0.8,0.4))
  else return(c(0.4,0.2))
}

##### FUNCIO USER
user <- function(user.info,day,prob.new){
  if(nrow(user.info) == 0){
    env$day.info[1,"DAU"] <- env$day.info[1,"DAU"] + 1
    env$day.info[1,"NU"] <- env$day.info[1,"NU"] + 1
    uid <- 1
    user.info[uid,] <-< profile(type="new",uid,day)
    env$uids <- append(env$uids,uid)
    return(uid)
  }
  else{
    p_new <- runif(1)
    if(p_new < prob.new){
      env$day.info[1,"DAU"] <- env$day.info[1,"DAU"] + 1
      env$day.info[1,"NU"] <- env$day.info[1,"NU"] + 1
      uid <- nrow(user.info)+1
      user.info[uid,] <-< profile(type="new",uid,day)
      env$uids <- append(env$uids,uid)
      return(uid)
    }
    else{
      n <- nrow(user.info)
      aux <- numeric(n)
      for(i in 1:n) aux[i] <- 1.001^i
      weights <- aux/sum(aux)
      uid <- sample(1:n,prob=weights,1)
      if(uid %in% env$uids == FALSE){
        if((day-1) == user.info[uid,"FirstPlay"]){
          env$day.info$R1 <- env$day.info$R1 + 1
        }
        if((day-7) == user.info[uid,"FirstPlay"]){
          env$day.info$R7 <- env$day.info$R7 + 1
        }
        if((day-30) == user.info[uid,"FirstPlay"]){

```

```

        env$day.info$R30 <- env$day.info$R30 + 1
    }
}
profile(type="existing",uid,day)
return(uid)
}
}
}

##### FUNCIO PROFILE (existing or new)
profile <- function(type="existing",uid,day){
  if(type == "existing"){
    user.info[uid,"TotalSessions"] <- as.numeric(
      user.info[uid,"TotalSessions"]) + 1
    if(uid %in% env$uids == FALSE){
      env$day.info[1,"DAU"] <- env$day.info[1,"DAU"] + 1
      user.info[uid,"DaysPlayed"] <- as.numeric(
        user.info[uid,"DaysPlayed"]) + 1
      env$uids <- append(env$uids,uid)
    }
  }
  if(type == "new"){
    country <- sample(countries,1)
    platform <- sample(platforms,1)
    clas <- behavior(type="new")
    return(c(uid,country,platform,"freeloader",day,clas,1,1,1,0))
  }
}

##### FUNCIO BEHAVIOR (probs or new)
behavior <- function(type="probs",clas){
  if(type == "probs"){
    if(clas == "spontaneous") return(c(0.1,0.01))
    if(clas == "casual") return(c(0.3,0.05))
    if(clas == "compulsive") return(c(0.7,0.1))
    if(clas == "whale") return(c(0.5,1))
  }
  if(type == "new"){
    p.clas <- runif(1)
    if(p.clas < 0.31){
      clas <- "spontaneous"
    } else if(p.clas < 0.72){
      clas <- "casual"
    } else if(p.clas < 0.98){
      clas <- "compulsive"
    } else clas <- "whale"
    return(clas)
  }
}

```



```

}
}

##### FUNCIO PLAY
play <- function(uid,prob.repeat){
  ngames <- rtpois(1,prob.repeat)
  user.info[uid,"Level"] <- as.numeric(
    user.info[uid,"Level"]) + ngames
  return(ngames)
}

##### FUNCIO ZERO-TRUNCATED POISSON (for ngames)
rtpois <- function(n,lambda,tol=1e-10){
  x <- rep(NA,n)
  low <- which(lambda < tol)
  nlow <- length(low)

  if(nlow > 0){
    x[low] <- 1
    if(nlow < n)
      x[-low] <- qpois(runif(n-nlow,dpois(
        0,lambda[-low]),1),lambda[-low])
  } else x <- qpois(runif(n-nlow,dpois(0,lambda),1),lambda)
  return(x)
}

##### FUNCIO PURCHASE
iap <- function(uid){
  if(user.info[uid,"Usertype"] == "freeloader"){
    user.info[uid,"Usertype"] <- "payer"
  }
  spent <- sample(prices,1,prob=prob.sales)
  user.info[uid,"TotalSales"] <- as.numeric(
    user.info[uid,"TotalSales"]) + spent
  env$day.info[1,"Revenue"] <- env$day.info[1,"Revenue"] + spent
  if(uid %in% env$payers == FALSE){
    env$day.info$Payers <- env$day.info$Payers + 1
    env$payers <- append(env$payers,uid)
  }
}
}

```

```
##### CARACTERISTIQUES DE LA SIMULACIO
tmax <- 86400
ndays <- 365
gametime <- 120
iaptime <- 10

countries <- c("XX","YY","ZZ")
platforms <- c("iOS","Android")

prices <- c(1.99,4.99,9.99,49.99,99.99)
prob.sales <- c(0.4,0.25,0.2,0.1,0.05)

##### VARIABLES
Uid <- integer()
Country <- factor(levels=countries)
Platform <- factor(levels=platforms)
Usertype <- factor(levels=c("freeloader","payer"))
FirstPlay <- integer()
Behavior <- factor(levels=
  c("spontaneous","casual","compulsive","whale"))
Level <- numeric()
DaysPlayed <- integer()
TotalSessions <- integer()
TotalSales <- numeric()
```